9th International Conference on Electrical and Computer Engineering
20-22 December, 2016, Dhaka, Bangladesh

A hybrid deep model with HOG features for Bangla handwritten numeral classification

S M A Sharif, Nabeel Mohammed, Nafees Mansoor, Sifat Momen
Department of Computer Science and Engineering
University of Liberal Arts Bangladesh (ULAB)
Satmasjid Road, Dhaka, Bangladesh
Email:{sma.sharif.cse,nabeel.mohammed,nafees.mansoor,sifat.momen}@ulab.edu.bd

***Abstract*—Considering the practical significances, handwriting recognition is getting an intense interest to the research community. Through, several studies have been conducted for Bengali handwriting recognition, a robust model for Bengali numerals classification is still due. Therefore, a hybrid model is presented in this paper, which aims to classify the Bengali numerals more precisely. The proposed model bridges hand crafted feature extraction based approaches with the automatically learnt features of Convolutional Neural networks (CNN). It is observed that the proposed model outperforms existing models with lesser epochs. The proposed model is trained and tested with the ISI numeral dataset and also cross-validated with the CAMTERDB numeral dataset. For both scenarios, proposed model shows consistency and demonstrate the maximum accuracy of 99.02% and 99.17%, respectively. For the CMATERDB collection, the proposed model achieves the best accuracy rate reported till date.**

I. Introduction

Due to the various application potentials, recognition of handwritten numerals has been receiving a profound interest to the computer vision researchers over the past few years. However, Bangla, the seventh most widely spoken language in the world and the second most used language in the Indian subcontinent, is lagging far behind in handwritten Bangla numeral research. Meanwhile, with progressive affordability of technology and incentives by the concerned Governments, Bangla document analysis is more relevant now than every before.

Deep learning methods, specifically convolutional neural network (CNN) based solutions exhibit notable performance in handwriting recognition, image classification, image annotation and various other fields. CNNs are multi-layered networks that learns, extracts and classifies features automatically. In contrast, other popular approaches such as SIFT [1],HOG [2] and LBP [3] are"hand crafted"', as these methods extract salient image information through techniques which were devised through experimentation and ingenuity of the researchers. Inspired by the performance of deep learning models a few recent studies have reported on the performance of CNN-based models for Bangla handwriting recognition.

The main contribution of this paper is the proposal of a hybrid model which combines the use of a automatically learnt CNN-based features and the hand crafted Histogram of Oriented Gradients (HOG) feature. The proposed model is trained on a data set of Bangla isolated numerals. It is observed that the proposed hybrid model outperforms other CNN-only methods in terms of accuracy and the number of epochs required to achieve the best accuracy. Moreover, through cross dataset testing, very good performance is observed by the proposed model even on untrained collections.

The rest of the paper is organized as follows. In §II, existing CNN-based Bangla numeral recognition methods are discussed. §III-D provides the details of the different experiments done for this research as well as the motivations. §IV presents the results that are obtained from the experiments. Next, conclusion and future works are presented in §V.

II. Background

In a breakthrough study [4] showed that CNNs can successfully be applied to large and complex image classifications tasks. Such CNNs are neural networks which accept gray scale or colour images as their input, and have successive layers of filters, which are the trainable weights, and pooling/subsampling layers [5]. Non-linearities, such as the sigmoid or ReLU functions, are typically applied after the convolutional layers. Such sequences of filtering and pooling layers are often followed by fully connected(FC) layers for classification purposes. These networks are effective because they are both feature learners and extractors.

Prior to the widespread adoption of deep learning, hand crafted features such as Histogram of Oriented Gradients (HOG) [2], Scale Invariant Feature Transform (SIFT) [1], Local Binary Patterns (LBP) [3] were among the most effective tools for image classification. These features were devised through experimentation and ingenuity of the researchers involved, and were considered to be state of the art. [6] published a detailed analysis of five different image features, their encoding schemes and classification performance on standard data sets. They identified HOG, which is based on image gradients being accumulated in localised histograms, as one of the most effective features. However, in a follow up paper, the authors compared the performance of such hand crafted approaches with deep learning techniques [7] and concluded that deep architectures outperform the hand crafted methods by a large margin [7]

A trend towards deep learning techniques can be observed in recent studies concerning Bangla handwriting recognition. [8] used a LeNet-like five layer CNN for character classification. This network has two sets of convolutional and subsampling layers in sequence, successively learning and extracting higher level features, with the last layer (F5) extracting a 300 dimensional feature vector, which is then used in fully connected layers during training. The motivation of this work was to train a single network on a single dataset and then re-use the network features. They initially trained their network on a data set with 50 classes (characters) and then applied the same network to other data sets, using the higher level features to train an SVM classifier. This method gave them a generalised

978-1-5090-2963-1/16/$31.00 ©2016 IEEE

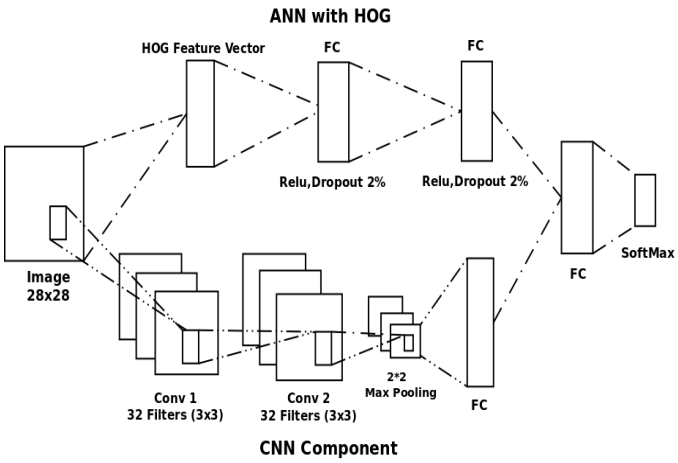


Fig. 1: Hybrid Model Overview

feature extractor (the CNN) tuned for characters which could then be used to train classifiers like SVM which do not demand the same volume of training data. Using this scheme they achieved 98.375% accuracy on the ISI Bangla numeral dataset. Recently [9] used a smaller CNN to train on the ISI numeral dataset and achieved more impressive results of 98.98% accuracy. In this case the network was specifically trained for the numeral dataset. As the training set is quite small ($19,392$ images), they augmented the training set by creating rotated versions of the training images. They trained different networks for rotation angles of $5°, 10°, 20°$ and $30°$. Best validation accuracy of 98.98% was achieved when the training set was augmented by $10°$ rotations. Unfortunately the merit of the augmentation method is questionable as for each experiment the images were rotated by a fixed angle, and then they reported the best observed results from multiple such experiments. This is equivalent to hand fitting for the testing set. Our attempts to reproduce the results on the same network, but when using random rotations, resulted in sub 98% validation accuracy.

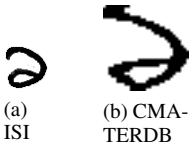
None of these studies have explored the efficacy of combining CNN's with handcrafted features for Bangla OHR. In this paper we present the use of a hybrid network which combines the use of features learnt and extracted by CNNs, with the very popular histogram of oriented gradients (HOG) image feature.

III. Experimental Setup

A. Proposed Hybrid model

The hybrid model proposed in this paper has two components, as shown in Figure 1. The first component is an Artificial Neural Network with two hidden layers of size 32 and 64 respectively. The input to this network is the HOG feature vector extracted from an image. The second component of the hybrid model is a traditional multi-layered CNN which has two convolutional layers in sequence, each with 32 3×3 filters, followed by a 2×2 max-pooling layer. The output of the sub-sampling layer is flattened to a vector and forwarded to a fully connected layer.

The output of the two components are combined into a larger feature vector, which then acts as the input to another network comprising of two fully connected layers, the last of which is a softmax layer which provides probability distributions of class predictions.



(a) ISI (b) CMATERDB

Fig. 2: Example images from the ISI Character database (a) and CMATERDB3.1.1(b)

The ANN component and the CNN component were also trained individually by forwarding their output to a 10 node softmax layer. We will be referring the the Hybrid model with the "Hybrid-" prefix. The CNN only and ANN only models will be refered by the "CNN" and "ANN-" prefixes.

B. Data Sets

This study utilises two independently collected data sets – the *Indian Statistical Institute (ISI) handwritten character database* [10], specifically the Bangla handwritten numeral data set, and the *CMATERDB3.1.1* collection [11]. Both data sets contain images of isolated handwritten Bangla numerals. The ISI images have variable size and contains noise. The CMATERDB3.1.1 images are all 32×32 pixels, mostly noise free but the character edges seem to be more blocky. Figure 2 shows an example image from each of the two collections.

For deep learning, the recommended approach is to have three partitions: a training set, a validation set and a testing set. The ISI collection comes with a partition of $19,392$ training images and $3,986$ testing images. The CMATERDB3.1.1 collection has 6000 images, which has been split into testing and training sets by different studies, usually 5000 training images and 1000 testing images. In this study, the ISI collection's training and testing partitions are used for training and validation respectively. The entire CMATERDB3.1.1 collection is used as the testing set.

C. Dataset preparation

A smoothing operation with a small kernel is performed at the very beginning to remove any noise from the images. The images are then converted to a binary format, followed by a *not* operation to have black background and white foreground. An edge thickening filter is applied next, and finally all images are resized to 28×28 pixels.

Following the recommendation of [7] each training image is rotated clockwise and anticlockwise by a random angle between $0° - 50°$, resulting in a final training set of $58,176$ training images. The difference with [9] is that rotation is performed by random angles, not fixed ones.

D. List of experiments and their motivation

Table I lists the different models trained to compare the performance of three different networks and their inputs described in §III-A. CNN-64 is very similar to the CNN described previously, except the first layer has 64 filters instead of 32. This network has been trained to observe the effect of increasing the number of network parameters, without introducing hand crafted features. A comparison between the performance of CNN and the Hybrid-* models can demonstrate the efficacy of including HOG features in the Hybrid model.

Experiment List			
Experiment Label	Network Type	Input	No. of Parameters
CNN	CNN	Grey scale 28×28 image	600810
CNN-64	CNN	Grey scale 28×28 image	610346
ANN-Hog-PPC4	ANN	HOG feature(392 dimensions)	15338
ANN-HOG-PPC8	ANN	HOG feature(72 dimensions)	5098
Hybrid-HOG-PPC4	Hybrid Model	HOG feature(392 dimensions) and 28×28 image	616138
Hybrid-HOG-PPC8	Hybrid Model	HOG feature(72 dimensions) and 28×28 image	605898

TABLE I: Details of the different models used in the experiments

E. Model training

All models were trained on the augmented training set of the ISI Bangla numeral collection. HOG features were extracted from all the images of the augmented set. The Adadelta optimiser was used with categorical cross entropy as the loss function. Each model was trained for 100 epochs.

IV. Results and Analysis

Fig. 3 illustrates how the loss values change over the training period, whereas Fig. 4 shows the how the accuracy values (training and validation) change over the epochs. Training results for different models are summarized in Table II, where the accuracies are evaluated over $3,986$ validating images. CNN, CNN-64 and Hybrid-HOG-PPC4 models achieve accuracy values around 98.9% in 100, 87 and 98 epochs, respectively. Higher number of parameters for the Hybrid-HOG-PPC4 is a result of the higher input dimensionality of the HOG feature vector (392). With lower number of parameters compared to Hybrid-HOG-PPC4 and CNN-64, Hybrid-HOG-PPC8 still shows better performance by achieving a validation accuracy of 99.02% in only 55 epochs.

Experiment Name	Accuracy(%)	Epoch
HOG-PPC4	96.61	98
HOG-PPC8	89.76	87
CNN	*98.87*	100
CNN-64	*98.87*	86
Hybrid-HOG-PPC4	98.90	98
Hybrid-HOG-PPC8	**99.02**	**55**

TABLE II: Best validation accuracy and the number of epochs required

Therefore, from the validation set results, it is observed that hand crafted feature that works well individually may not be suitable with learnt CNN features (Hybrid-HOG-PPC4). Moreover, it is also observed that a well-chosen hand crafted feature can improve the accuracy of a deep model and the training speed.

Performances of the proposed Hybrid-HOG-PPC8 and other existing models are presented in Table III for ISI Bangla numeral data set and CMATERDB3.1.3 data set. Table III shows that Hybrid- Hog-PPC8 achieves an accuracy of 99.17% on the CMATERDB3.1.1 data set, where a collection of 6000 images is considered for the test. Till date, this is the most precise recognition rate for the CMATERDB3.1.3 collection. Moreover, it is worth to mention that the 6000 test images are not at all used in the training phase, which explains the robustness of the proposed model.

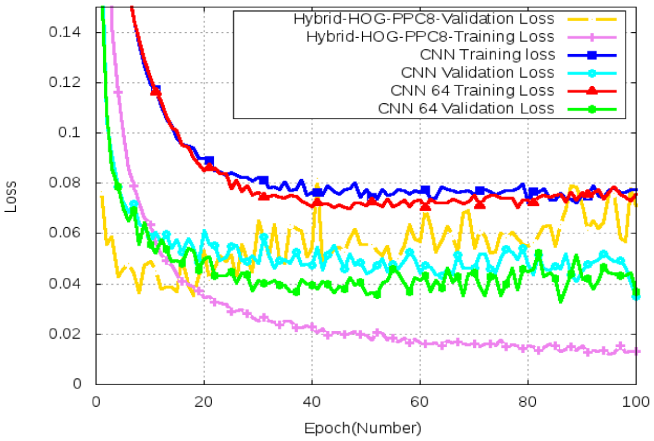


Fig. 3: Loss values over the training period

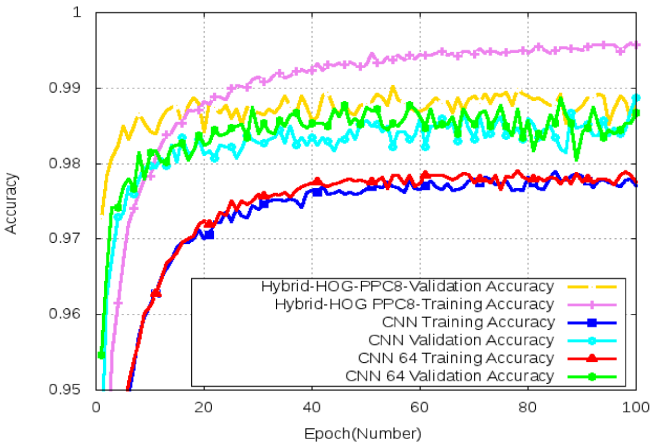


Fig. 4: Accuracy values over the training period)

V. Conclusion

This paper presents a hybrid deep learning model that combines CNN with HOG features. Bangla handwritten isolated numeral data set is considered to train the proposed model. From the experimental results, it is observed that the proposed hybrid model achieves better accuracy with lesser epochs as compared to other CNN only models. Moreover, in cross data set testing, the proposed model exhibits the most precise recognition rate for handwritten Bengali numerals. Moreover, the proposed model outperforms CNNs with greater number of parameters, thereby demonstrating the efficacy of using the HOG features. In future, the proposed hybrid model will be integrated with multiple handcrafted features and behavior of that particular system will be studied.

ISI Numeral Dataset		CMATERDB 3.1.1	
Work	Accuracy	Work	Accuracy
Bhattacharya and Chaudhuri [10]	98.20%	Haider Adnan Khan et al. [12]	94%
Wen and He [13]	96.91%	Hassan et al. [14]	96.7%
Das et al. [15]	97.70%	Das et al. [16]	98.55%
Nasir and Uddin [17]	96.80%	Sarkhel el at. [18]	98.23%
Akhnad et al. [19]	97.93%	Basu et al. [20]	97.15%
CNNAP, [9]	98.98 %	Basu et al. [21]	95.1%
Hybrid-HOG-PPC8 (proposed)	**99.02%**	Hybrid-HOG-PPC8 (Proposed)	**99.17%**

TABLE III: Performance comparison of Hybrid-HOG-PPC8 with previously published results

REFERENCES

[1] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[2] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1. IEEE, 2005, pp. 886–893.

[3] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105. [Online]. Available: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

[5] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995.

[6] K. Chatfield, V. S. Lempitsky, A. Vedaldi, and A. Zisserman, "The devil is in the details: an evaluation of recent feature encoding methods." in *BMVC*, vol. 2, no. 4, 2011, p. 8.

[7] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," *arXiv preprint arXiv:1405.3531*, 2014.

[8] D. S. Maitra, U. Bhattacharya, and S. K. Parui, "Cnn based common approach to handwritten character recognition of multiple scripts," in *Document Analysis and Recognition (ICDAR), 2015 13th International Conference on*. IEEE, 2015, pp. 1021–1025.

[9] M. M. H. R. M. A. H. Akhand, Mahtab Ahmad, "Convolutional neural network training with artificial pattern for bangla handwritten numeral recognition," *ICIEB*, vol. 1, no. 1, pp. 1–6, 2016.

[10] U. Bhattacharya and B. B. Chaudhuri, "Handwritten numeral databases of indian scripts and multistage recognition of mixed numerals," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 3, pp. 444–457, 2009.

[11] R. Sarkar, N. Das, S. Basu, M. Kundu, M. Nasipuri, and D. K. Basu, "Cmaterdb1: a database of unconstrained handwritten bangla and bangla–english mixed script document image," *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 15, no. 1, pp. 71–83, 2012.

[12] H. A. Khan, A. Al Helal, and K. I. Ahmed, "Handwritten bangla digit recognition using sparse representation classifier," in *Informatics, Electronics & Vision (ICIEV), 2014 International Conference on*. IEEE, 2014, pp. 1–6.

[13] Y. Wen and L. He, "A classifier for bangla handwritten numeral recognition," *Expert Systems with Applications*, vol. 39, no. 1, pp. 948–953, 2012.

[14] T. Hassan and H. A. Khan, "Handwritten bangla numeral recognition using local binary pattern," in *Electrical Engineering and Information Communication Technology (ICEEICT), 2015 International Conference on*. IEEE, 2015, pp. 1–4.

[15] N. Das, R. Sarkar, S. Basu, M. Kundu, M. Nasipuri, and D. K. Basu, "A genetic algorithm based region sampling for selection of local features in handwritten digit recognition application," *Applied Soft Computing*, vol. 12, no. 5, pp. 1592–1606, 2012.

[16] N. Das, J. M. Reddy, R. Sarkar, S. Basu, M. Kundu, M. Nasipuri, and D. K. Basu, "A statistical–topological feature combination for recognition of handwritten numerals," *Applied Soft Computing*, vol. 12, no. 8, pp. 2486–2495, 2012.

[17] M. K. Nasir and M. S. Uddin, "Hand written bangla numerals recognition for automated postal system," *IOSR Journal of Computer Engineering*, vol. 8, no. 6, pp. 43–48, 2013.

[18] R. Sarkhel, N. Das, A. K. Saha, and M. Nasipuri, "A multi-objective approach towards cost effective isolated handwritten bangla character and digit recognition," *Pattern Recognition*, vol. 58, pp. 172–189, 2016.

[19] S. I. P. S. Md. Mahbubar Rahman, M. A. H. Akhand and M. M. H. Rahman, "Bangla handwrittencharacter recognition using convolutional neural network," *I.J.Image, Graphics and Signal Processing(IJIGSP*, vol. 7, no. 3, pp. 42–49, 2015.

[20] S. Basu, N. Das, R. Sarkar, M. Kundu, M. Nasipuri, and D. K. Basu, "A novel framework for automatic sorting of postal documents with multi-script address blocks," *Pattern Recognition*, vol. 43, no. 10, pp. 3507–3521, 2010.

[21] S. Basu, R. Sarkar, N. Das, M. Kundu, M. Nasipuri, and D. K. Basu, "Handwritten bangla digit recognition using classifier combination through ds technique," in *Pattern Recognition and Machine Intelligence*. Springer, 2005, pp. 236–241.